DOCID: 4233816

~~TOP SECRET~~

TICOM/I-89

REPORT BY PROF. DR. H. ROHRBACH

OF PERS. ZS, ON

AMERICAN STRIP CYPHER.

Attached is an account prepared by the above P.O.W. on the cryptographic handling of the American Strip Cypher C-2.

The homework was done at Marburg on August 6, 1945 and was brought back by Major W. Bundy, A.U.S.

Ticom

14th Sept. 1945

No. of pages: 14

DISTRIBUTION

British

D.D.3
D.D.4
D.D. (N.S.)
D.D. (M.W.)
D.D. (A.S.)
Lt. Col. Leathem
Cdr. Tandy
Major Morgan
Miss Mortimer (2)

U.S.

Op-20-G (2) (via Lt. Cdr. Manson)
G-2 (via Lt. Col. Hilles)
A.S.A. (3) (via Major Seaman)
Director, S.I.D. USFET
(via Lt. Col. Johnson)
Col. Lewis Powell, USSTAF

TICOM

Chairman
S.A.C. (2)
Cdr. Bacon
Lt. Col. Johnson
Major Seaman
Lt. Cdr. Manson
Capt. Cowan
Lt. Fehl
Ticom Files (2)

Do NOT Destroy. Return to the NSA Technical Library when no longer needed
S-243
Copy No.

Approved for Release by NSA on 08-25-2015. FOIA Case # 74892

DOCID: 4233816

Marburg/Lahn, August 6, 1945
23 Lutherstrasse

[To Chairman, TICOM.

This was probably of no value. It was assigned to appease vanity and keep people busy.

W.P. BUNDY.]

Sir,

I enclose herewith the report as commissioned on the decipherment of the American strip cipher O-2 by the German Foreign Office.

The production was handicapped by the inadequate working conditions owing to our present accommodation. Therefore we confined ourselves to strip cipher O-2 only. The way we worked this one out shows our typical methods for breaking a strip cipher if sufficient material is available.

Furthermore, we were not in a position to give all the care we wanted to the drawing up of the report. I can not guarantee for absolute correctness of dates etc., since we had to rely solely on our memory, there being no files at our disposal.

So I ask you to overlook kindly any internal or external defects of this report. Further details are at your disposal if desired.

Faithfully yours,

(signed) Prof. Dr. H. Rohrbach

REPORT ON

THE DECIPHERMENT OF THE AMERICAN STRIP CIPHER O-2 BY THE GERMAN FOREIGN OFFICE.

MARBURG 1945

Contents

A. Introduction
1. Personnel
2. Selection of material.

B. Sorting out the material
3. Preliminary investigation
4. Parallels
5. Working hypothesis
6. Formation of classes
7. Lists of blocks

C. Construction of complexes out of nuclei
8. Nuclei and complexes
9. Classification of nuclei
10. Production of nuclei
11. Comparison of nuclei
12. Further processing

DECLASSIFIED 4233816

Contents (continued)

D. From complex to family

13. Depth and correlation of bigrams
14. Extension of complexes (Methods)
15. Extension of complexes (Execution)
16. Refining of complexes
17. Supplementary review

E. Philological solution

18. Trial
19. Solution

F. Establishment of strips and keys

20. The first 30 strips
21. The remaining 20 strips
22. The keys for the day

G. The decipherment of the material

23. The automaton

The American Strip Cipher O-2

A. Introduction

1. Personnel. The present report on the American Strip Cipher O-2 describes the systematic solution of this cipher by the special Service (Pers Z S) of the German Foreign Office. The necessary work was carried out by the mathematical-cryptographic section under the direction of ORR Kunze Ph. D. His scientific collaborators in this task had been prepared by several year's special training for decipher work. They were Mrs. Annelise Hünke Ph.D., Miss Erika Pannwitz Ph.D., Assistant Professor Helmut Grunsky Ph.D., Studienrat Hansgeorg Krug, Professor Hans Rohrbach Ph.D., and Mr. Klaus Schultz. Towards the end of the work these were joined by Mr. Hans-Kurt Mueller Ph.D. of the linguistic-cryptographic section as a philologist.

In drawing up this report, which was written jointly by the above collaborators, only those steps could be mentioned that eventually led to the solution. As a matter of fact, many detours had to be made and, through numerous individual attempts, observations and experience had to be gathered which could not be gone into here. Besides an expense of time of more than a year, the strictly enforced principle that the whole staff of collaborators should take part in all the processes of the work was decisive for its success.

These cryptographers had a large number of trained assistants at their disposal. In order to work and to take full advantage of the very ample traffic Hollerith-machines of different types were employed to a large extent. Finally, after the conclusion of the work, all messages were deciphered by means of a special machine (the automaton) invented by Mr. Kunze.

2. Selection of material. In sorting out the American diplomatic messages currently intercepted in 1940 and 1941 the share of telegrams enciphered with still unknown ciphers grew more and more. From the peculiarities of the cipher texts it was at first assumed that they had been enciphered by means of a cryptographic machine. But steadily continued investigations showed that the number of keys could not be so large as might be expected in the case of modern cryptographic machines. In the summer of 1942 the number of messages enciphered with the new cipher had become so big that the possibility of

DOCID: 4233816

deciphering them even in case of a complicated cipher had to be taken into account. Therefore, in November 1942 systematic work was begun with a view to breaking a certain part of the material. This consisted of the messages of the intermission traffic (marked by five-letter date groups ending in y or i), and it was selected because here the number of identical beginnings of the cipher text and the size of the material were especially large.

B. Sorting out the material

3. Preliminary investigation. The first 40 letters of the cipher text of all inter-mission messages together with some technical data (date, traffic, number of message etc.) were punched on Hollerith cards and sorted out from different points of view in lexicographic order. Out of these sorted cards all those pairs of cards were selected mechanically where at least five consecutive letters in corresponding places coincided. The result was written down on lists. The following facts were observed:

(1) A large number of parallels (repetitions of groups of consecutive letters) appeared, partly of considerable length, some extending up to the first 30 letters.

(2) No parallel extended beyond the 30th letter.

(3) An unusually high percentage of parallels broke off with the 15th letter.

(4) There were some parallels beginning with the 16th letter only, while the preceding 15 letters were free from doublets, i.e. differed from each other pairwise.

(5) Parallels occurring in messages of different days were never of the same month, often the time interval between them was several months.

(6) Parallels were especially frequent in messages of the same day.

(7) There was no parallel between messages before August 1, 1942 and messages after that date.

From these observations could be deducted that always 15 letters of the text were enciphered in the same way [see (3) and (4)], in some cases even 30 letters [see (1) and (2)]. Further, the encipherment had to be dependent on the date [see (5) and (6)], any day key being capable of being used several times. Finally on August 1, 1942 approximately there must have been a change of the means of encipherment [see (7)].

4. Parallels. On the assumption that even after August 1, 1942, the same means of encipherment would be valid for a length of time, the material coming in after that date was worked upon. Each message was divided into double lines of 30 letters each. In each double line a left-hand line and a right-hand line of 15 letters each were distinguished. For each double line a Hollerith card was punched indicating text, date, traffic as well as a four-figure number. With a daily average of 15 messages of an average length of 20 double lines 50.000 punched cards were on hand at the beginning of the work. At a later stage of the work the months January to March 1943 were added, so that about 80.000 punched cards in all were available. From these cards all parallels of at least five letters each were sorted out and tabulated by means of Hollerith machines. Although with the immense size of the material a large number of the five-letter-parallels and partly also of the six-letter-parallels were bound to be accidental i.e. false, they were still added in the hope that the shorter ones of these parallels would confirm each other and that thus even the days with few messages could be secured.

DOCUMENT 4233816

The result of that tabulation confirmed all of the above-mentioned observations (1) to (7) and the conclusions drawn from them. Beyond that it was observed that:

(8) All parallels of at least 8 letters appeared vertically below each other, i.e. began in the same place of their respective lines.

(9) In some cases four to five different days of the period in question belonged to a certain day key. The total of the messages belonging to one and the same key was called class.

(10) Within a class longer parallels were also to be found in the interior of the messages. From the frequency of their occurrence it could be concluded that only relatively few possibilities of encipherment were available for the lines of a class. No law regulating the change of encipherment of the lines could be recognized. All of the lines of a class enciphered in the same way were termed family.

(11) A family consisted of either only left-hand lines or right-hand lines.

(12) Beginning about March 1, 1943, only the messages of the (very numerous) Berne traffic and of some minor missions had parallels, so that the remaining traffic had to be eliminated from the material of March 1943.

5. Working hypothesis. From the outline of the cipher text and from the occurrence of frequent initial parallels whose true reading could in all probability be considered known (e.g. strictly confidential, from Murphy) judging from previously broken codes it could be deduced that each letter of the true reading corresponded to exactly one letter, different from it, of the cipher text. Since moreover, genuine parallels were always located exactly one below the other, the assumption was obvious that the lines of a family were enciphered with a column cipher, i.e. that all first letters were enciphered with their own substitution alphabet, all second letters with a second alphabet etc., so that 25 substitution alphabets had to be established for the solution of a family.

On the other hand it was known that so-called strip ciphers were being used in the American non-diplomatic code service. The following peculiarities are characteristic for such a cipher:

(a) There are 25 different reading distances.

(b) Each reading distance produces one substitution alphabet for each strip.

(c) If the same true reading is read at two different distances the two cipher texts are free from doublets.

Peculiarity (b) coincided with our assumption that we dealt with a column cipher. Peculiarity (c) corresponded to our observation (4) [see No. 3] that there were parallels which were preceded or followed at the ends of the lines by longer passages free from doublets. Thus it was fairly certain and in the course of the work it became an absolute certainty that the cipher in question was a strip cipher. To each reading distance corresponded a family and vice versa, so that every class (same day key) comprised 25 families in both halves.

6. Formation of classes. Now the direction in which the work had to proceed was prescribed: first, to find out the strongest class, then to crystallise one or several of the strongest families out of the strongest class.

Through a careful study of the above parallels, furnished by Hollerith machines, and considering cyclic correlations (by demanding that two days well correlated to a third should be well correlated to each other), we succeeded to sort about two-thirds of the days into approximately 44 classes. This number could still diminish through the growing together of classes. As a matter of fact, there were

DOCID: 4233816

40 different classes after the solution. The days which could not be inserted into this class calendar had mostly short or few messages, so that in fact about 80% of the total material was divided into classes.

The strongest of these classes (class III) contained ca. 3.000 lines. The following days belonged to it: Sept. 17, 1942; Oct. 16, 1942; Nov. 27, 1942; Jan. 30, 1943, and Mar. 19, 1943. Hereinafter only this class and its left-hand families [see No. 4, (11)] will be dealt with. We were faced with the task of constructing families out of its lines. There being 25 left-hand families, each family consisted of 120 lines on an average or - having regard to the dispersion - of about 60 to 180 lines.

7. Lists of blocks. As the next step thorough experiments were made with synthetic cipher texts. Experience was gathered along the following lines: to what extent have any two lines of 15 letters each to coincide in parallels, triplets of letters (trigrams), pairs of letters (Bigrams) and single letters, so that, with reasonable safety, they may be regarded as belonging to the same family. Not being able to count on finding all the lines of a family, we confined ourselves, at the beginning of the work, to those lines which had at least one bigram in common.

The next object was therefore, to copy the 3.000 left-hand lines of the chosen class arranged according to bigrams. Since every line contains 14 bigrams, it was bound to occur in this list in 14 different places; the total list thus contains 14 x 3.000 = 42.000 lines. The list began with the lines containing the bigram aa as first and second letters, followed by those containing ab, ac, ... az, ba, ...zz. Then followed the same for the second bigrams, (i.e. bigrams in the second and third place of the lines) and so on up to the fourteenth bigrams (in the fourteenth and fifteenth places). As the lines having a bigram in common were separated on the list from those of the following bigram by a blank line, they formed little blocks for the eye (averaging 4 to 5 lines), so that this list was called list of blocks.

This work could, of course, be carried out only with the aid of Hollerith machines; the punched cards of the selected class were already on hand and needed only to be sorted out from the total material. As several collaborators were to be charged with the utilization of the lists of blocks, they were typed in multiplicate.

C. Construction of complexes out of nuclei.

8. Nuclei and complexes. A complex was understood to be an aggregate of lines (of a breadth of 15 letters each) presumably belonging to one and the same family on account of frequent coincidences of letters located below each other (good correlations). The number of lines was called the depth of the complex.

A first attempt at constructing larger complexes directly by joining isolated lines, proved a failure. This procedure was recognised as unreliable by way of a trial on a model and rejected.

First of all, 2-line-nuclei and 3-line-nuclei (i.e. complexes of 2 or 3 lines resp. with especially good correlations) were produced as a starting point for the formation of complexes. The following types of nuclei were used:

DIGITAL 4233816

(1) 2-line-nuclei with correlation [4,1] and better, i.e. with coinciding letters in four consecutive columns and at least one additional column.

(2) 3-line-nuclei with a through-bigram and additional cyclic correlation [2,2,1] and better, i.e. all three lines coincided in two consecutive columns; of the three pairs of lines which could be formed from a 3-line-nucleus, two were to be correlated in at least two additional columns and the third in at least one additional column by coinciding letters - apart from the through-bigram.

9. Classification of nuclei. The 2-line-nuclei and 3-line-nuclei were classified according to quality from the following points of view:

(1) The larger the number of correlations the better the nucleus.
(2) Continuous correlations were better than dispersed ones; e.g. in the case of 2-line-nuclei [6] was better than [5,1] or [4,2]; a 3-line-nucleus with a through-trigram was the total of correlations being equal - better than one with a through-bigram.
(3) When in a 3-line-nucleus with through-bigram the additional correlation - e.g. [2,2,1] - resulted from the fact that yet another coincidence continued through all three lines, the nucleus was valued less than one in which the correlation of the several pairs of lines were all located in different columns (exception: through-trigram, see above).

Thus a 3-line-nucleus was generally the better the fewer letters it contained, if coinciding letters in each column were counted only once.

The valuation of nuclei was based on research on mathematical probability concerning the expectation about the different correlation types in accidental (i.e. not belonging to one and the same family) pairs of lines and triplets of lines as well as investigations of the true reading on the occurrence of parallels. The later work, too, was currently supplemented by purely mathematical research, the very extensive calculations of which were laid down in numerous tables and nomograms. This research could not be gone into further within the scope of this report.

10. Production of nuclei. The 2-line-nuclei, insomuch as they contained a five-letter-parallel, had already been ascertained during the search for parallels (see No. 4). In order to find the remaining 2-line-nuclei and 3-line-nuclei, in the 14 lists of blocks (see No. 7) within each block:

(1) The parallels immediately following the through-bigram were underlined.
(2) Letters occurring repeatedly in each column were encircled.
(3) The nuclei of the desired type of correlation were sorted out.

Of these three operations only the third required qualified assistants. The result was about 50 2-line-nuclei and 360 3-line-nuclei, a considerable portion of which, according to research on mathematical probability had to be considered as genuine.

11. Comparison of nuclei. In order to join several nuclei to complexes, they had to be compared pair by pair, and those pairs of nuclei had to be sorted out which showed a specially large number of coincidences in corresponding columns. For this purpose each nucleus was punched into a Hollerith card after a special method. When punched cards of two nuclei were put on top of each other, two holes coincided if and only if they represented the same letter in corresponding columns of the two nuclei - apart from the steering holes in the upper margin of the punching area. The number of through-holes in the pair of cards thus equalled the number of correlations of the pair of nuclei through coinciding letters.

DOCID: 4233816

With the aid of punched cards it was possible to have 300 to 400 nuclei compared with a certain nucleus by untrained assistants within one hour.

Pairs of 3-line-nuclei were sorted out if they contained at least 8 correlations, pairs consisting of a 3-line-nucleus and a 2-line-nucleus, if they contained at least 6 correlations. (Expected values in the case of an average of 36 and 25 resp. different letters in the columns of 3-line-nuclei and 2-line-nuclei resp.: 4 and 3 correlations resp.)

The result of the joining of 2-line- and 3-line-nuclei were 4-line-, 5-line-, and 6-line-nuclei.

12. Further processing. As had been done for the 2-line- and 3-line-nuclei, Hollerith-cards were punched for the new 4-line-, 5-line-, and 6-line-nuclei. The new nuclei were compared with each other as well as with the 2-line- and 3-line-nuclei. In this way the depth of the nuclei was increased successively. In the case of nuclei of the depth of 10 or more the procedure was modified insofar as no longer all the letters of the nucleus but only the frequent letters of a column were punched into the card.

The mechanical comparison of nuclei by means of punched cards could only serve to find suitable building stones for the construction of larger complexes out of the great quantity of material.

In any case before the joining of a pair of nuclei to a bigger nucleus the corresponding lines were written down and tested by an expert as to the quality of the nuclei, the quality of the correlations of each line in the whole complex, a possible decomposition into several components, the number of the different letters, and resemblance to a true reading (parallels etc.). A large number of nuclei eliminated itself during the course of the work, because an attempt at obtaining greater depth yielded only vague results.

The fusion of nuclei resulted in about 40 complexes of a depth of about 20 lines each. For further processing they were marked with capitals A, B, ..., AA, BB, ...

There being only 25 families, one had to take into account that some of those 40 complexes either belonged to one and the same family or that a considerable portion of the complexes was so badly constructed, that they could not be coordinated to any family. There were indeed pairs of complexes which presumably belonged to the same family, since - in spite of all efforts to keep the complexes separated - ever and again there were found nuclei which showed good correlations with the two complexes of the pair. One of those pairs consisted of the complexes E and K.

D. From complex to family

13. Depth and correlation of bigrams. Among the complexes arising from the fusion of 3-line-nuclei there was a particularly good one. It had been named E and included 21 lines; it was called E 21. A measure for the quality of a complex was obtained in the following way: The 15 columns of a complex of the depth t were split up into 14 pairs of columns, in each of these pairs of columns the number of different bigrams was ascertained and the

DOCID: 4233816

sum of all these numbers divided by 14; the result was a value t_b, which was called the depth with reference to bigrams or in short bigram depth of the complex. The quotient t_b/t gave a measure for the quality of the complex, and the complex was the better, the smaller the value of t_b/t.

The next step consisted in studying the complexes A, B, C.... whether some of them - in analogy to the 3-line-nuclei - were capable of growing together into a bigger one. As preparatory work for this a true reading experiment was made in order to obtain a criterion for the minimum number of bigram correlations necessary for the fusion of two complexes. A bigram correlation was said to exist, if in a pair of columns the same bigram occurred in the one as well as in the other complex. It appeared (see conclusion of No. 12) that the complex E_{21} grew together with a complex K_{16} ($t = 16$) to a complex of the depth $t = 37$; it was named E_{37}. The value t_b/t for E_{37} when compared with that for E_{21} showed a slight deterioration which suggested the existence of false lines (i.e. not belonging to the complex).

14. Extension of complexes (Methods). The next task was to rise gradually from the complex E_{37} to the corresponding family through the addition of new lines from class III. The family of any (pure) complex C was called the family of C, and two methods were developed by which additional lines to C could be found from class III.

(a) Bigram method. For each pair of columns of C the bigrams were counted statistically and compared with the corresponding bigrams of all remaining lines of class III. In this way a bigram statistic was obtained, which furnished the bigram correlations with C for every line of class III not belonging to C. From a special nomogram could be read how many lines with a prescribed number of bigram correlations might be expected accidentally for a complex of the considered bigram depth, and thus a new possibility of testing the quality of C was obtained. The lines with the most bigram correlations came into question as candidates for the family of C.

(b) Weight method. In each column of C the frequency of each letter a, b, ..., z was counted, then - for every column separately - these sums were arranged according to size and the letters divided into three groups (frequent, medium, rare). Frequent letters received the weight 2, medium ones the weight 1, rare ones the weight 0. Finally in every line of class III we wrote above each one of the 15 letters its appropriate weight (which in general was different for different columns) and by adding the 15 individual weights the total weight of the line was ascertained. The lines of highest weight (so far as they did not belong to C already) were also candidates for the C family, but coincided partly with the lines obtained by method (a).

By erecting a pillar for each (line-) weight - representing the number of lines from class III with this weight - on a scale running from left to right, a graphic representation of the distribution of weights in the form of a bell-shaped Gauss curve was obtained. However, on the right-hand side of the chief maximum and distinctly separated from it, there arose above the highest weights another, smaller maximum, which derived from the lines of class III belonging to the C family. By marking the lines of C specially (e.g. colouring them) one could practically read from the second maximum how many lines from class III could still be added. The

DOCID: 4233816

second maximum was the more marked, the purer the underlying complex, i.e. the fewer false lines it contained. Thus the graphic representation gave an additional possibility of testing the quality of a complex.

15. Extension of complexes (Execution). By applying and combining these two methods (a) and (b) repeatedly E_{37} was extended to E_{51}, thence to E_{95}, then to E_{120} and finally to E_{135}. The selection of lines to be added from the candidates obtained by methods (a) or (b) was a task to be undertaken with particular care. Only rarely the bigram correlations or the maximum weights were valued by themselves. In general both factors were decisive jointly. Moreover, the quantity of lines that had been found in this way was investigated as to whether corroborations occurred through bigram correlations of these lines among themselves. Where corroboration was lacking lines were excluded if need be. Soon, however, doubts arose as to whether it was permissible to apply methods (a) and (b) for adding new lines any number of times. For one had to take into account that wrong lines might be added with every step, which might involve additional wrong lines with the next step and thus might deteriorate the complex more and more. A procedure was devised therefore to eliminate wrong lines as far as possible.

16. Refining of complexes. This procedure was as follows:

(c) Method of reproduction. The complex E_{95} was split into two parts E_I and E_{II}. E_I was identical with E_{51}; E_{II} consisted of the 44 lines that had been added when E_{51} was extended to E_{95}. Then methods (a) and (b) were applied to E_{II}. Had all lines of E_{II} been genuine, E_I should have been reproduced substantially out of it. This, however, was not the case. Besides many other lines of class III only 26 lines of E_I were reproduced. Now these 26 lines were considered as a new starting complex E_{26} and methods (a) and (b) were applied to it again step by step. Thus we proceeded, even more scrupulously than before, from E_{26} to E_{50}, thence to E_{75} and finally to E_{89}. To eliminate wrong lines the weight method was applied after each step even in a refined form, viz. the letters of each column were divided into five instead of three classes, with the weights of 4, 3, 2, 1, 0 according to frequency. Again lines that were found wanting on thorough examination were eliminated.

The complex E_{89} thus obtained was handed over to a philologist for further treatment. At the same time, for safety, E_{89} just as previously E_{95} was treated according to method (c). It was split into two parts $E_{III} = E_{50}$ and $E_{IV} = E_{89} - E_{50}$, and the attempt was made to reproduce E_{III} out of E_{IV}. This time the attempt was much more successful. 43 lines were obtained again out of E_{50}. These lines were taken as a new starting complex E_{43} and extended to a complex E_{82} by methods (a) and (b). It had been planned to treat this complex - which could with great probability be regarded as pure - with the same methods (a), (b) and (c), in order to extend it to a depth of about 120 lines while preserving its purity as far as possible. But this proved unnecessary as in the meantime the philologist had already succeeded in solving complex E_{89}.

DOCID: 4233816

17. Supplementary Review. The key valid for class III having been completely solved, all lines were deciphered subsequently. In that way all the families could be established and the quality of the complexes worked on could be ascertained. Complex E_{95} proved to contain about 31 (30%) wrong lines, complex E_{89} eight (9%), and complex E_{82} two (2.5%). The corresponding family included ca. 145 lines altogether.

Apart from the complexes E_n other complexes were treated in a similar way. Great difficulties arose, however, in the construction of further complexes with sufficient depth. The methods applied had forcibly led to the most favourable family, i.e. the family with the best correlations. Probability considerations had shown that such a family could be counted upon. For the discovery of this family, however, it had been necessary to plan the preparatory work described in sections B and C on as wide a basis as had been done.

E. Philological solution.

18. Trial. The usual frequencies of letters and combinations of letters in the English language could not be applied to this complex unrestrictedly: first on account of the peculiar style of the telegrams, then because the complex was composed of 89 small pieces from different telegrams without any inherent connection, and finally, because we dealt only with well-correlated lines. Therefore we began with a trial solution on a model complex. From several true readings taken from messages encoded with the Grey and Brown Codes, we chose arbitrarily 100 lines of 15 letters each and wrote them one below the other. This gave 15 columns with a depth of 100 letters each. These 15 columns were enciphered with 15 different substitution alphabets. The solution yielded certain deviations of frequencies from those of the usual English language. A striking peak was formed by the bigram -in-. The combination -tion was noticeably frequent as well. On the other hand -th- belonged to the less frequent bigrams, contrary to the ratio of frequency in the usual English language.

19. Solution. The text of E_{89} - the appropriate family to be called E for short - contained the frequent combination qqjw in the last 4 columns, twice even the combination pqqjw. The interpretations -tion and -ation were obvious. This hypothesis furnished one interpreted letter for each of the last 4 columns. In the other columns as well striking repetitions of letter-combinations were to be found, above all frequencies of bigrams. The bigram -in- was substituted in several pairs of columns by way of an experiment.

Additional help came from another side. The family E* reciprocal to E (i.e. the coordination of true reading elements to cipher elements in E* is equal to the coordination of cipher elements to true reading elements in E) was made use of. It was known that the E* family contained among its lines the initial line of a message, which could easily be interpreted. Inverting the resulting correspondences of letters gave one letter - more or less rare - for each column of E_{89}.

The study of bigram frequencies in the text E_{89} was continued carefully and they were compared with those of the model mentioned above. E.g. if in one pair of columns the bigram xy occurred several times, the bigram vy twice and the bigram wy three times, the interpretations -in, -an, -on or -in, -un, -an were obvious.

DOCID: 4233816

After the greatest possible number of bigrams had been interpreted and substituted in this systematic way, parts of words began here and there to shine forth from the text. Filling the gaps furnished new values of letters, and the complete solution progressed practically of its own accord, once this stage had been reached.

F. Establishment of strips and keys

20. The first 30 strips. After Eg9 had been solved, i.e. after the true reading - though not completely at first - had been established for the 81 genuine lines, we received a substitution alphabet, still containing gaps, for each of the 15 lines. This alphabet made possible the coordination of cipher-text letters to true reading letters for the different columns. In order to complete the alphabets the true reading had to be established for as many more lines as possible. This was arrived at in two ways. On the one hand all the lines of the material were deciphered with these substitution alphabets. In this way all those lines of the E family were found which had not been discovered by the statistical methods previously applied; that gave about 65 additional lines. On the other hand we sorted out the lines of the E* family from the total material of class III. This E* family comprised approximately 60 lines. In this way all gaps in the 15 alphabets could be filled. With these complete alphabets the first 15 strips were constructed on the arbitrary assumption that in the E family coordinated cipher elements and clear elements were located side by side on the strips. Since on this assumption the letters on the strips could be arranged in a single - i.e. not decomposable - cycle, this arbitrary arrangement corresponded to an odd power of the original arrangement, as it was on the original strips used for the encipherment of the texts. Later on, when the original arrangement had been discovered through systematic studies of the - psychologically conditioned - cipher habits of the code clerks with the different missions, it became evident that it was the eleventh power of the original arrangement.

By means of the 15 finished strips all of the left-hand lines of class III could be completely deciphered. The resulting true reading could be continued into the right-hand lines in many places, so that there the coordinations of cipher elements to clear elements - necessary for the construction of further strips - were given. In this way the 15 strips employed for the encipherment of the right-hand lines were obtained step by step. And thereby all 30 strips necessary for the decipherment of class III were available.

21. The remaining 20 strips. The next task was to find out whether and to what extent these 30 known strips appeared in other classes as well. This task was essentially simplified by the fact that after studying the true readings in class III the true reading in characteristic places could - in all probability - be considered as known for other classes as well. Were for a definite class in one line.

$g_1 g_2 g_3 \ldots$ cipher elements, $k_1 k_2 k_3 \ldots$ the corresponding clear elements,

in a second line

$g'_1 g'_2 g'_3 \ldots$ cipher elements, $k'_1 k'_2 k'_3 \ldots$ the corresponding clear elements

DOCID: 4233816

in a third line $g_1''g_2''g_3''$ ... cipher elements, $k_1''k_2''k_3''$ ... the corresponding clear elements a.s.o.

and were $S_1S_2S_3$... the strips causing these coordinations, there had to exist fixed numbers a, a', a", ..., so that on the strip

S_1 the distance $g_1 - k_1 = a$, $g_1' - k_1' = a'$, $g_1'' - k_1'' = a''$, ...

S_2 the distance $g_2 - k_2 = a$, $g_2' - k_2' = a'$, $g_2'' - k_2'' = a''$, ...

S_3 the distance $g_3 - k_3 = a$, $g_3' - k_3' = a'$, $g_3'' - k_3'' = a''$, ...

The next task was to verify whether any of the strips already available answered the conditions mentioned above for $S_1S_2S_3$... When class I, ranking next in size to class III was investigated in this way, 18 of its strips turned out to be already known from class III.

22. The keys for the day. Thereby the problem of the total number of the different strips was solved: since 12 strips were added to 18 known ones 20 more were to be expected in addition to the 30 known ones. This was confirmed when all different classes one after the other were treated according to the above-mentioned method. In the course of this treatment the unknown strips appearing in individual columns could be constructed by supplementing the true readings in these columns. As soon as all of the 50 strips and all of the 40 day keys were finished, the calendar could be completed according to which the day keys were apportioned to the individual days. Herewith all messages enciphered with the system 0-2 could be read.

G. The decipherment of the material.

23. The automaton. As was to be foreseen at the outset the total material could not be deciphered by hand on account of its immense size. The number of available qualified workers with sufficient knowledge of English was too small for that. Deciphering a double line through moving the strips by hand required 6 - 7 minutes on an average, so that the work on the roughly 80.000 double lines would have taken a whole year, provided that 4 collaborators had worked on it 8 hours daily. It was, therefore, of the utmost importance that the automaton should be available for the decipherment of the material at the time when all keys had been worked out. It is not possible to describe the machine more explicitly within the scope of this report, but we should like to say briefly the following about the method of its working:

The decipherment of a double line consists of two operations: (1) arranging the strips so that the cipher text letters are made to lie in a row, (2) selecting the line containing the true reading out of 25 parallel lines. The adjustment of the strips that move up and down, so that the true reading can be read horizontally, is accomplished by the machine quite automatically. The cipher text may be touched by hand on the keyboard of a typewriter, or be taken by means of a sounding device from the Hollerith cards that had already been punched. Finding the true reading is simplified by the fact that the letters on the strips are printed in two different weights, the most frequent letters in the English language (about 80% of true reading) are printed in a heavy tone, the others in a light tone. A line consisting of 15 letters chosen arbitrarily would contain

DOCID: 4233816

6 bold ones on an average, while the true reading line of 15 letters with 12 bold ones on an average stands out distinctly; moreover this line is indicated by a bright spot of light on the edge of the set of strips. The 30 strips necessary for the decipherment of a double line are arranged side by side in two groups of 15 each for the line; if the left-hand group is in the first movement, the right-hand one is in the second movement and vice versa. During the time when the clerk copies the true reading from the indicated line on the typewriter, the machine prepares automatically the adjustment of strips for the next line and performs it at the touch of a key. In this way the decipherment of a double line requires barely half a minute on an average. By means of this machine the total material could be deciphered within a month.